

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/87980 A2(51) International Patent Classification⁷: C07K 14/705, C12N 15/12, C07K 16/28, A01K 67/027, A61K 38/17

Gude Drive, Rockville, MD 20850 (US). CRAVCHIK, Anibal; Celera, 45 West Gude Drive, Rockville, MD 20850 (US). DI FRANCESCO, Valentina; Celera, 45 West Gude Drive, Rockville, MD 20850 (US). BEASLEY, Ellen M.; Celera, 45 West Gude Drive, Rockville, MD 20850 (US).

(21) International Application Number: PCT/US01/15957

(74) Agent: CELERA GENOMICS CORP.; Robert A. Millman, Patent Director, 45 West Gude Drive C2-4, Rockville, MD 20850 (US).

(22) International Filing Date: 17 May 2001 (17.05.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

(30) Priority Data:
60/205,196 18 May 2000 (18.05.2000) US
09/634,656 8 August 2000 (08.08.2000) US

(71) Applicant: APPLERA CORPORATION [US/US]; Robert A. Millman, Assistant Secretary, 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: WEI, Ming-Hui; Celera, 45 West Gude Drive, Rockville, MD 20850 (US). ZHAO, Qi; Celera, 45 West

[Continued on next page]

(54) Title: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

```

1 ATGAATGAGC CACTAGACTA TTTAGCAAAT GCTTCTGATT TCCCCGATTA
51 TGCAGCTGCT TTTGGAAATT GCACATGATGA AAACATCCCCA CTCAAGATGC
101 ACTACCTCCC TGTTATTTAT GGCATTATCT TCTCTGTTGG ATTTCAGGCG
151 AATGCAGTAG TGATATCCAC TTACATTTTC AAAATGAGAC CTTGGAAAGAG
201 CAGCACCATC ATTATGCTGA ACCTCTGCCCT CACAGATCTG CTGTATCTGA
251 CCAGCCCTCCC CTTCCTGATT CACTACTATG CGACGTGGCA AAACATGGATC
301 TTGGAGATT TCATGTGTA GTTTATCCGG TTCAAGCTTCC ATTTCACCTT
351 GTATAGCAGC ATTCCTCTTCC TCACCCCTGGT CAGCATCTTC CGCTACTCTG
401 TGATCATTCA CCCCATGAGC TGCTTTTCCA TTACACAAAC TCAGATGTGCA
451 GTTGTAGCCCT GTGCTGTGGT GTGGATCATT TCACCTGGTAG CTGTCATTCC
501 GATGACCTTC TTGATCACAT CAACCAACAG GACCAACAGA TCAGCCCTGTC
551 TCGACCTTCAG CGATTCGGAT GAACTCAATA CTATTAAGTG GTACAACCTG
601 ATTTTGACTG CAACTACTTT CTGCTCTCCC TTGGTGATAG TGACACTTGG
651 CTATPACACG ATTATCCACA CTCTGACCCCA TTGACTGGCAA ACTGACAGCT
701 GCCTTAAGCA GAAAGCACGA AGGGCTAACCA TTCTGCTACT CCTTGGATT
751 TACGTTATGTT TTTTACCTTT CCATATCTTG AGGGTCATTC GGATCGAATC
801 TCGGCTGCTT CTAATCAGT TCTCTTCTGAA GTATCAGATC CATGAACTT
851 ACATCTGTTTC TAGACCTTCAG GCTGCTCTGA ACACCTTTGG TAACCTGTGTA
901 CTATATGTGG TGTCACCGGA CAACCTTCAG CAGGGCTGTCT GTCTCAACAGT
951 GAGATGCAAA GTAAAGGGGA ACCTTGAGCA ACAGAAAGAAA ATTAGTTACT
1001 CAAACAAACCC TTGA

```

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the Human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

FEATURES:
Start: 1
Stop: 1012

HOMOLOGOUS PROTEIN:

gi 6679193 ref NP_032798.1 purinergic receptor P2Y, G-protein ...	226	2e-58
gi 450557 ref NP_002554.1 purinergic receptor P2Y, G-protein ...	223	2e-57
gi 1352695 sp P49651 P2YR_RAT P2Y PURINOCEPTOR 1 (ATP RECEPTOR)...	222	3e-57
gi 464327 sp P34996 P2YR_CHICK P2Y PURINOCEPTOR 1 (ATP RECEPTOR...)	221	7e-57
gi 1352693 sp P49652 P2YR_MELGA P2Y PURINOCEPTOR 1 (ATP RECEPTO...)	221	7e-57
gi 1352691 sp P48042 P2YR_BOVIN P2Y PURINOCEPTOR 1 (ATP RECEPTO...)	221	7e-57
gi 1082756 pir SS4253 purinergic receptor - human >gi 798836 e...	217	1e-55
gi 2829680 sp P79928 P2Y8_XENLA P2Y PURINOCEPTOR 8 (P2Y8) >gi 1...	217	2e-55
gi 2707256 gb AAC60339.1 (AF031897) G protein coupled P2Y nucl...	215	7e-55

blast to dbEST:
gi|7921097|gb|AW827323.1|AW827323 hm31f01.x1 NCI_CGAP_Thy4 Homo... 46 0.022
gi|2900654|gb|AA826657.1|AA826657 of34e08.s1 NCI_CGAP_Kid6 Homo... 42 0.35Expression information for modulatory use (from cDNA panel screening):
Human uterus
Human Placenta
Human Brain
Human Fetal Brain

WO 01/87980 A2



patent (AM, AZ, BY, KG, KZ, MD, RU, T1, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *without international search report and to be republished upon receipt of that report*

**ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID
MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF**

RELATED APPLICATIONS

5 The present application claims priority to U.S. Serial No. 09/634,656, filed August 8, 2000 (Atty. Docket CL000750) and U.S. Serial No. 60/205,196, filed May 18, 2000 (Atty. Docket CL000567-PROV).

FIELD OF THE INVENTION

The present invention is in the field of G-Protein coupled receptors (GPCRs) that are related 10 to the purinergic receptor subfamily, recombinant DNA molecules and protein production. The present invention specifically provides novel GPCR peptides and proteins and nucleic acid molecules encoding such protein molecules, for use in the development of human therapeutics and human therapeutic development.

15 **BACKGROUND OF THE INVENTION**

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three 20 extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular 25 inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest.* 92:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet.* 30:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet.* 26:403-424(1992)), and nephrogenic diabetes 30 insipidus (Holtzman *et al.*, *Hum. Mol. Genet.* 2:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary

analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the $\beta 2$ -purinergic receptor and currently represented by over 200 unique members
5 (Dohlman *et al.*, *Annu. Rev. Biochem.* 60:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science* 254:1024-1026 (1991); Lin *et al.*, *Science* 254:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science* 258 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science* 241:1467-1472 (1988)); and Family V, the fungal 10 mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem.* 61:1097-1129 (1992)).

There are also a small number of other proteins that present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein that has been extensively studied and does not show evidence of being 15 a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5047-5051 (1993)). The gene *frizzled* (fz) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, fz has not been shown to couple to G-proteins (Vinson *et al.*, *Nature* 338:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α , β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors 20 containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α -subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the $\beta\gamma$ -subunits. The GTP-bound form of the α -subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenyl cyclase), diacylglycerol or inositol phosphates. Greater 25 than 20 different types of α -subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed 30 in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

Purinergic GPCRsPurinoceptor P2Y1

P2 purinoceptors have been broadly classified as P2X receptors which are ATP-gated channels; P2Y receptors, a family of G protein-coupled receptors, and P2Z receptors, which mediate 5 nonselective pores in mast cells. Numerous subtypes have been identified for each of the P2 receptor classes. P2Y receptors are characterized by their selective responsiveness towards ATP and its analogs. Some respond also to UTP. Based on the recommendation for nomenclature of P2 purinoceptors, the P2Y purinoceptors were numbered in the order of cloning. P2Y1, P2Y2 and P2Y3 have been cloned from a variety of species. P2Y1 responds to both ADP and ATP. Analysis of P2Y 10 receptor subtype expression in human bone and 2 osteoblastic cell lines by RT-PCR showed that all known human P2Y receptor subtypes were expressed: P2Y1, P2Y2, P2Y4, P2Y6, and P2Y7 (Maier et al. 1997). In contrast, analysis of brain-derived cell lines suggested that a selective expression of P2Y receptor subtypes occurs in brain tissue.

Leon et al. generated P2Y1-null mice to define the physiologic role of the P2Y1 receptor. (J. 15 Clin. Invest. 104: 1731-1737(1999)) These mice were viable with no apparent abnormalities affecting their development, survival, reproduction, or morphology of platelets, and the platelet count in these animals was identical to that of wildtype mice. However, platelets from P2Y1-deficient mice were unable to aggregate in response to usual concentrations of ADP and displayed impaired aggregation to other agonists, while high concentrations of ADP induced platelet aggregation without shape change. 20 In addition, ADP-induced inhibition of adenylyl cyclase still occurred, demonstrating the existence of an ADP receptor distinct from P2Y1. P2Y1-null mice had no spontaneous bleeding tendency but were resistant to thromboembolism induced by intravenous injection of ADP or collagen and adrenaline. Hence, the P2Y1 receptor plays an essential role in thrombotic states and represents a potential target 25 for antithrombotic drugs. Somers et al. mapped the P2RY1 gene between flanking markers D3S1279 and D3S1280 at a position 173 to 174 cM from the most telomeric markers on the short arm of chromosome 3. (Genomics 44: 127-130 (1997)).

Purinoceptor P2Y2

The chloride ion secretory pathway that is defective in cystic fibrosis (CF) can be bypassed by an alternative pathway for chloride ion transport that is activated by extracellular nucleotides. 30 Accordingly, the P2 receptor that mediates this effect is a therapeutic target for improving chloride secretion in CF patients. Parr et al. reported the sequence and functional expression of a cDNA cloned from human airway epithelial cells that encodes a protein with properties of a P2Y nucleotide receptor. (Proc. Nat. Acad. Sci. 91: 3275-3279 (1994)) The human P2RY2 gene was mapped to chromosome 11q13.5-q14.1.

Purinoceptor P2RY4

The P2RY4 receptor appears to be activated specifically by UTP and UDP, but not by ATP and ADP. Activation of this uridine nucleotide receptor resulted in increased inositol phosphate formation and calcium mobilization. The UNR gene is located on chromosome Xq13.

5

Purinoceptor P2Y6

Somers et al. mapped the P2RY6 gene to 11q13.5, between polymorphic markers D11S1314 and D11S916, and P2RY2 maps within less than 4 cM of P2RY6. (Genomics 44: 127-130 (1997)) This was the first chromosomal clustering of this gene family to be described.

Adenine and uridine nucleotides, in addition to their well established role in intracellular 10 energy metabolism, phosphorylation, and nucleic acid synthesis, also are important extracellular signaling molecules. P2Y metabotropic receptors are GPCRs that mediate the effects of extracellular nucleotides to regulate a wide variety of physiological processes. At least ten subfamilies of P2Y receptors have been identified. These receptor subfamilies differ greatly in their sequences and in their nucleotide agonist selectivities and efficacies.

15 It has been demonstrated that the P2Y1 receptors are strongly expressed in the brain, but the P2Y2, P2Y4 and P2Y6 receptors are also present. The localisation of one or more of these subtypes on neurons, on glia cells, on brain vasculature or on ventricle ependimal cells was found by in situ mRNA hybridisation and studies on those cells in culture. The P2Y1 receptors are prominent on neurons. The coupling of certain P2Y receptor subtypes to N-type Ca²⁺ channels or to particular K⁺ channels was 20 also demonstrated.

25 It has also been demonstrated that several P2Y receptors mediate potent growth stimulatory effects on smooth muscle cells by stimulating intracellular pathways including Gq-proteins, protein kinase C and tyrosine phosphorylation, leading to increased immediate early gene expression, cell number, DNA and protein synthesis. It has been further demonstrated that P2Y regulation plays a mitogenic role in response to the development of atherosclerosis.

It has further been demonstrated that P2Y receptors play a critical role in cystic fibrosis. The volume and composition of the liquid that lines the airway surface is modulated by active transport of ions across the airway epithelium. This in turn is regulated both by autonomic agonists acting on basolateral receptors and by agonists acting on luminal receptors. Specifically, extracellular 30 nucleotides present in the airway surface liquid act on luminal P2Y receptors to control both Cl⁻ secretion and Na⁺ absorption. Since nucleotides are released in a regulated manner from airway epithelial cells, it is likely that their control over airway ion transport forms part of an autocrine regulatory system localised to the luminal surface of airway epithelia. In addition to this physiological

role, P2Y receptor agonists have the potential to be of crucial benefit in the treatment of CF, a disorder of epithelial ion transport. The airways of people with CF have defective Cl⁻ secretion and abnormally high rates of Na⁺ absorption. Since P2Y receptor agonists can regulate both these ion transport pathways they have the potential to pharmacologically bypass the ion transport defects in CF.

5 GPCRs, particularly members of the purinergic receptor subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing a previously unidentified human GPCR.

10 **SUMMARY OF THE INVENTION**

The present invention is based in part on the identification of nucleic acid sequences that encode amino acid sequences of human GPCR peptides and proteins that are related to the purinergic subfamily, allelic variants thereof and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for 15 the development of human therapeutic targets, aid in the identification of therapeutic proteins and serve as targets for the development of human therapeutic agents.

The proteins of the present inventions are GPCRs that participate in signaling pathways mediated by the purinergic subfamily in cells that express these proteins (see expression information in Figure 1, the GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as 20 determined by cDNA panel screening). As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR protein. Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP₂), inositol 1,4,5-triphosphate (IP₃) and adenylate cyclase; polarization of the plasma membrane; production or secretion 25 of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival

The response mediated by the receptor protein depends on the type of cell it is expressed on. Some information regarding the types of cells that express other members of the subfamily of GPCRs of the present invention is already known in the art (see references cited in Background and 30 information regarding closest homologous protein provided in Figure 2 and expression information provided in Figure 1, the GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP

metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the particular GPCR of the present invention, a skilled artisan will clearly know that the receptor protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., 5 through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell thus participating in a biological process in the cells or tissues that express the GPCR (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP₂) as well as to 10 the activities of these molecules. PIP₂ is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP₂ to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP₃). Once formed IP₃ can diffuse to the endoplasmic reticulum surface where it can bind an IP₃ receptor, e.g., a calcium channel protein containing an IP₃ binding site. IP₃ binding can 15 induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP₃ can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP₄), a molecule that can cause calcium entry into the cytoplasm from the extracellular medium. IP₃ and IP₄ can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP₂) and inositol 1,3,4- 20 triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP₂. The other second messenger produced by the hydrolysis of PIP₂, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as 25 the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF- κ B. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP₂ or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved 30 in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This

activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane
5 depolarization.

By targeting an agent to modulate a GPCR, the signaling activity and biological process mediated by the receptor can be agonized or antagonized in specific cells and tissues (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). Such agonism and antagonism serves as a basis for modulating a
10 biological activity in a therapeutic context (mammalian therapy) or toxic context (anti-cell therapy, e.g. anti-cancer agent).

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule or transcript sequence
15 that encodes the GPCR of the present invention. In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

20 FIGURE 2 provides the predicted amino acid sequence of the GPCR of the present invention. In addition structure and functional information, such as protein family and function, modification sites, is provided that allows one to readily determine specific uses of inventions based on this molecular sequence as well as significant fragments of the proteins of the present invention.

FIGURE 3 provides genomic sequences that span the gene encoding the GPCR protein of
25 the present invention. In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided that allows one to readily determine specific uses of inventions based on this molecular sequence as well as important fragments for use in probe and primer design and heterologous gene expression control. Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following
30 variations were seen: A6613T, A7338G and A7732C.

DETAILED DESCRIPTION OF THE INVENTION**General Description**

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed 5 previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a GPCR protein or part of a GPCR protein, that are related to the purinergic subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention 10 provides amino acid sequences of human GPCR peptides and proteins, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these GPCR peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the GPCR of the present invention.

15 In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known GPCR proteins of the purinergic subfamily and the expression pattern observed (The GPCR of the present invention is expressed in the brain, uterus, 20 placenta and fetal brain as determined by cDNA panel screening.). The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of 25 the known purinergic family or subfamily of GPCR proteins.

Specific Embodiments**Peptide Molecules**

The present invention provides nucleic acid sequences that encode protein molecules that 30 have been identified as being members of the GPCR family of proteins (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious

variants described herein, such as allelic variants, will be referred herein as the GPCR peptides of the present invention, GPCR peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of or are comprised of the amino acid sequences of the GPCR peptides disclosed 5 in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA and Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of 10 cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide 15 having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes 20 preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the GPCR peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% 25 chemical precursors or other chemicals.

The isolated GPCR peptide can be purified from cells that naturally express it, purified from 30 cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). For example, a nucleic acid molecule encoding the GPCR peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequences that such a protein consists of is provided in Figure 2. A 5 protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in 10 Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that are comprised of the amino acid sequences 15 provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein is comprised of an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded 20 sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the GPCR peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

25 The GPCR peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a GPCR peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the GPCR peptide. "Operatively linked" indicates that the GPCR peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the GPCR 30 peptide.

In some uses, the fusion protein does not affect the activity of the GPCR peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of

recombinant GPCR peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A GPCR peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCR peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art know techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the GPCR peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at

that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

5 The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; 10 *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at 15 <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 20 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. 25 The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to 30 obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic*

Acids Res. 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence 5 identity to one of the GPCR peptides of the present invention as well as being encoded by the same genetic locus as the GPCR peptide provided herein. The GPCR of the present invention is encoded by a gene on chromosome 13 near markers SHGC-68276 (LOD=8.71) and SHGC-8118 (LOD=7.82).

Allelic variants of a GPCR peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the GPCR peptide as 10 well as being encoded by the same genetic locus as the GPCR peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human (The GPCR of the present invention is encoded by a gene on chromosome 13 near markers SHGC-68276 (LOD=8.71) and SHGC-8118 (LOD=7.82).). As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid 15 sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides SNP information that has been found in the gene encoding the GPCR 20 proteins of the present invention. The following variations were seen: A6613T, A7338G and A7732C.

Paralogs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide, as being encoded by a gene 25 from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a GPCR peptide can readily be identified as having some degree of significant 30 sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under

moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the GPCR peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the GPCR peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a GPCR peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant GPCR peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to bind G-protein, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis that identifies critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as ligand/effectector molecule binding or in assays such as an *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the GPCR peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly fragments identified in Figure 2.

The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16 or more contiguous amino acid residues from a GPCR peptide. Such fragments can be chosen based on the ability to retain one or 5 more of the biological activities of the GPCR peptide or could be chosen for the ability to perform a function, e.g. ability to bind ligand or effector molecule or act as an immunogen. Particularly important fragments are biologically active fragments, peptides which are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the GPCR peptide, e.g., active site, a G-protein binding site, a transmembrane domain or a ligand-binding 10 domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well-known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

15 Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in GPCR peptides are described in basic texts, detailed monographs, and the 20 research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent 25 attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, proteolytic processing, phosphorylation, prenylation, 30 racemization, selenylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Accordingly, the GPCR peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature GPCR peptide is fused with another compound, such as a compound to increase the half-life of the GPCR peptide (for example, polyethylene glycol),

or in which the additional amino acids are fused to the mature GPCR peptide, such as a leader or secretory sequence or a sequence for purification of the mature GPCR peptide or a pro-protein sequence.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures and Back Ground Section; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or receptor) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the binding partner so as to develop a system to identify inhibitors of the binding interaction. Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, GPCRs isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug to modulate the cells or tissues that express the receptor (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). Approximately 70% of all pharmaceutical agents modulate

the activity of a GPCR. A combination of the invertebrate and mammalian ortholog can be used in selective screening methods to find agents specific for invertebrates. The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention. Such uses can readily be determined using the information 5 provided herein, that known in the art and routine experimentation.

The receptor polypeptides (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions that are specific for the subfamily of GPCRs that the one of the present invention 10 belongs to, particularly in cells and tissues that express this receptor (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

The receptor polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the receptor protein, as a 15 biopsy or expanded in cell culture (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). In one embodiment, however, cell-based assays involve recombinant host cells expressing the receptor protein.

The polypeptides can be used to identify compounds that modulate receptor activity of the protein in its natural state, or an altered form that causes a specific disease or pathology associated with 20 the receptor. Both the GPCRs of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Further, these compounds can be tested in animal or invertebrate 25 systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree.

Further, the receptor polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a molecule that normally interacts with the receptor protein, e.g. a ligand or a component of the signal pathway that the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol 30 turnover and/or adenylate cyclase, or phospholipase C activation). Such assays typically include the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such

as G-protein phosphorylation, cAMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 5 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 10 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that 15 competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal 20 transduction pathway that indicate receptor activity. Thus, a cellular process such as proliferation, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade, can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase.

Any of the biological or biochemical functions mediated by the receptor can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described 25 herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the receptor can be assayed (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

30 Binding and/or activating compounds can also be screened by using chimeric receptor proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can be used that

interacts with a different G-protein than that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or 5 subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test 10 compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The receptor polypeptides are also useful in competition binding assays in methods designed to 15 discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the receptor target. This type of assay is particularly useful in cases in which compounds are sought that 20 interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

25 Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., 35 S-labeled) and the candidate compound, 30 and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from

the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein 5 trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor 10 protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the GPCRs of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems 15 are well known in the art and can readily be employed in this context.

Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells or tissues that express the GPCR (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). These methods of treatment include the 20 steps of administering a modulator of the GPCR's activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the GPCR proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) 25 *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the GPCR and are involved 30 in GPCR activity. Such GPCR-binding proteins are also likely to be involved in the propagation of signals by the GPCR proteins or GPCR targets as, for example, downstream elements of a GPCR-mediated signaling pathway. Alternatively, such GPCR-binding proteins are likely to be GPCR inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a GPCR protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other

construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a GPCR-dependent complex, the DNA-binding and activation domains of the transcription factor are 5 brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the GPCR protein.

10 This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a GPCR modulating agent, an antisense GPCR nucleic acid molecule, a GPCR-specific antibody, or a GPCR-binding partner) can be used in an animal or other model to determine the 15 efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or insect model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

20 The GPCR proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected. 25 Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

30 The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion,

insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques 5 useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagents, such as an antibody or protein binding agent.. Alternatively, the peptide can be detected *in vivo* in a 10 subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

15 The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic 20 failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or 25 therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the 30 receptor protein in which one or more of the receptor functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to

maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein (The GPCR of the present invention is expressed 5 in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). Accordingly, methods for treatment include the use of the GPCR protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present 10 invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. 15 In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal 20 antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is 25 administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods.

30 Antibodies are preferably prepared from regions or discrete fragments of the GPCR proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or receptor/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

5 Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline 10 phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable 15 radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the 20 purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and 25 pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

30 Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function, particularly in cells and tissues that express the receptor (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of

expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

10 Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical 15 assays known to those in the art.

15 The antibodies are also useful for tissue typing (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

20 The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the GPCR peptide to a binding partner such as a ligand. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against 25 intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

25 The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

5 The present invention further provides isolated nucleic acid molecules that encode a GPCR peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the GPCR peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

10 As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected 15 to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

20 Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

25 For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

30 Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figures 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ

ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule, for example from about 1-300 additional nucleotides.

The present invention further provides nucleic acid molecules that are comprised of the 5 nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule is comprised of a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a 10 fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these 15 nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the 15 source of the present invention, human genomic sequences (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be 20 identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity 25 modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or 25 carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular 30 enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the GPCR peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional

non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates

5 purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding

10 strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the GPCR proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs

15 (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-

20 conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify

25 gene modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing

30 regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene (The GPCR of the present invention is encoded by a gene on chromosome 13 near markers SHGC-68276 (LOD=8.71) and SHGC-8118 (LOD=7.82).).

Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: A6613T, A7338G and A7732C.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2 (Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: A6613T, A7338G and A7732C.).

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

5 The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into 10 the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

15 The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods (The GPCR of the present invention is encoded by a gene on chromosome 13 near markers SHGC-68276 (LOD=8.71) and SHGC-8118 (LOD=7.82).). This is particularly useful in determining whether a particular protein is an allelic variant of one the proteins provided herein

20 The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention as described in detail below.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

25 The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

30 The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to

assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in GPCR protein expression relative to normal results.

5 *In vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

10 Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a GPCR protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

15 Nucleic acid expression assays are useful for drug screening to identify compounds that modulate GPCR nucleic acid expression, particularly in cells and tissues that express the receptor (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

20 The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the GPCR gene. The method typically includes assaying the ability of the compound to modulate the expression of the GPCR nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired GPCR nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the GPCR nucleic acid (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.) or recombinant cells genetically engineered to express specific nucleic acid sequences.

25 The assay for GPCR nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the GPCR protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

30 Thus, modulators of GPCR gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of GPCR mRNA in the presence of the candidate compound is compared to the level of expression of GPCR mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of

mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

5 The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate GPCR nucleic acid expression, particularly to modulate activities within a cell or tissue that expresses the proteins (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). Modulation includes both up-regulation (i.e. activation or agonization) or 10 down-regulation (suppression or antagonization) of nucleic acid expression.

15 Alternatively, a modulator for GPCR nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the GPCR nucleic acid expression in the cells and tissues that express the protein (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.).

20 The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the GPCR gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the 25 affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

25 The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in GPCR nucleic acid, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in GPCR genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally-occurring 30 genetic mutations in the GPCR gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the GPCR gene associated with a dysfunction

provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a GPCR protein.

Individuals carrying mutations in the GPCR gene can be detected at the nucleic acid level by a variety of techniques (The GPCR of the present invention is encoded by a gene on chromosome 13 near markers SHGC-68276 (LOD=8.71) and SHGC-8118 (LOD=7.82).). Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis (Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: A6613T, A7338G and A7732C.). RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a GPCR gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant GPCR gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the GPCR gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: A6613T, A7338G and A7732C.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control GPCR gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of GPCR protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into GPCR protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of GPCR nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired GPCR nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the GPCR protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in GPCR gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired GPCR protein to treat the individual.

5 The invention also encompasses kits for detecting the presence of a GPCR nucleic acid in a biological sample, particularly cells and tissues that normally express the protein (The GPCR of the present invention is expressed in the brain, uterus, placenta and fetal brain as determined by cDNA panel screening.). For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting GPCR nucleic acid in a biological sample; means for determining 10 the amount of GPCR nucleic acid in the sample; and means for comparing the amount of GPCR nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCR protein mRNA or DNA.

15 Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or 20 oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in 25 their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in 30 length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length

of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically
5 examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for
10 one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

15 In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum
20 system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

25 In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the
30 microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be

obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, expression patterns, 5 mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the GPCR proteins/peptides of the present invention and allelic variation within this gene/protein. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such 10 assays will typically involve arrays comprising many genes or alleles, at least one of which is a gene and or alleles of the GPCR gene of the present invention (Figure 3 provides SNP information that has been found in the gene encoding the GPCR proteins of the present invention. The following variations were seen: A6613T, A7338G and A7732C.).

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation 15 conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein.

Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and 20 Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

25 The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

30 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the GPCR disclosed herein; and (b) one or

more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified GPCR genes of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

15

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are 20 covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate 25 into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

30 Expression vectors contain *cis*-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a *trans*-acting factor interacting with the *cis*-regulatory control region to allow transcription of the

nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

5 The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

10 In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

15 In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*.
20 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40,
25 Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

30 The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme 5 digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and 10 CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for 15 affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E 20 binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant 25 protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in 30 yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYEPSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured

insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian 5 expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for 10 maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., *Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein 15 are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, 20 tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein 25 into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., *Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be 30 introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one

vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be 5 replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers 10 include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation 15 systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as GPCRs, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to 20 these peptides.

Where the peptide is not secreted into the medium, which is typically the case with GPCRs, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid 25 extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the 30 cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a GPCR protein or peptide that can be further purified to produce desired amounts of GPCR protein or fragments. Thus, host cells containing expression 5 vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the GPCR protein or GPCR protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native GPCR protein is useful for assaying compounds that stimulate or inhibit GPCR protein function.

10 Host cells are also useful for identifying GPCR protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant GPCR protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native GPCR protein.

15 Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic 20 animal. These animals are useful for studying the function of a GPCR protein and identifying and evaluating modulators of GPCR protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a 25 pseudopregnant female foster animal. Any of the GPCR protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct 30 expression of the GPCR protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic 5 animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected 10 systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* *PNAS* 89:6232-6236 (1992). Another example of a recombinase system 15 is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.* *Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be 20 provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according 25 to the methods described in Wilmut, I. *et al.* *Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant 30 female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, GPCR protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* GPCR protein function, including ligand interaction, the effect of specific mutant GPCR proteins on GPCR protein function and ligand interaction, and the effect of chimeric GPCR proteins. It is also possible to assess

the effect of null mutations, that is mutations that substantially or completely eliminate one or more GPCR protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the

5 invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or

10 related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence shown in SEQ ID NO:2;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
 - (c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and
 - (d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.
3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

- (a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;
- (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);
- (d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and
- (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.
8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
9. A host cell containing the vector of claim 8.
10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.
13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.
14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.
16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human proteases, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human protease peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.

22. An isolated nucleic acid molecule encoding a human protease peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

1/7

```

1  ATGAATGAGC  CACTAGACTA  TTTAGCAAAT  GCTTCTGATT  TCCCCGATTA
51  TGCAGCTGCT  TTTGGAAATT  GCACTGATGA  AAACATCCCA  CTCAGAGATGC
101 ACTACCTCCC  TGTTATTTAT  GGCATTATCT  TCCTCGTGGG  ATTCCAGGC
151 AATGCAGTAG  TGATATCCAC  TTACATTTTC  AAAATGAGAC  CTTGGAAGAG
201 CAGCACCATC  ATTATGCTGA  ACCTGGCTG  CACAGATCTG  CTGTATCTGA
251 CCAGCCTCCC  CTTCCCTGATT  CACTACTATG  CCAGTGGCGA  AAACCTGGATC
301 TTTGGAGATT  TCATGTGTA  GTTTATCCGC  TTCAGCTTCC  ATTCAACCT
351 GTATAGCAGC  ATCCTCTTCC  TCACCTGTT  CAGCATCTTC  CGCTACTGTG
401 TGATCATTCA  CCCAATGAGC  TGCTTTCCA  TTCACAAAAC  TCGATGTGCA
451 GTTGTAGCCT  GTGCTGTGGT  GTGGATCATT  TCACTGGTAG  CTGTCATTCC
501 GATGACCTTC  TTGATCACAT  CAACCAACAG  GACCAACAGA  TCAGCCTGTC
551 TCGACCTCAC  CAGTTGGAT  GAACTCAATA  CTATTAAGTG  GTACAACCTG
601 ATTTGACTG  CAACTACTTT  CTGCTCCCC  TTGGTGATAG  TGACACTTTG
651 CTATACCACG  ATTATCCACA  CTCTGACCCA  TGGACTGCAA  ACTGACAGCT
701 GCCTTAAGCA  GAAAGCACGA  AGGCTAACCA  TTCTGCTACT  CCTTGCATTT
751 TACGTATGTT  TTTTACCCCT  CCATATCTTG  AGGGTCATTC  GGATCGAATC
801 TCGCCTGCTT  TCAATCAGTT  GTTCCATTGA  GAATCAGATC  CATGAAGCTT
851 ACATCGTTTC  TAGACCATT  GCTGCTCTGA  ACACCTTTGG  TAACCTGTTA
901 CTATATGTGG  TGGTCAGCGA  CAACTTCAG  CAGGCIGTCT  GCTCAACAGT
951 GAGATGCAA  GTAAAGCGGGG  ACCTTGAGCA  AGCAAAGAAA  ATTAGTTACT
1001 CAAACAACCC  TTGA

```

FEATURES:

Start: 1
 Stop: 1012

HOMOLOGOUS PROTEIN:

gi 6679193 ref NP_032798.1 purinergic receptor P2Y, G-protein ...	226	2e-58
gi 4505557 ref NP_002554.1 purinergic receptor P2Y, G-protein ...	223	2e-57
gi 1352695 sp P49651 P2YR_RAT P2Y PURINOCEPTOR 1 (ATP RECEPTOR)...	222	3e-57
gi 464327 sp P34996 P2YR_CHICK P2Y PURINOCEPTOR 1 (ATP RECEPTOR...	221	7e-57
gi 1352693 sp P49652 P2YR_MELGA P2Y PURINOCEPTOR 1 (ATP RECEPTOR...	221	7e-57
gi 1352691 sp P48042 P2YR_BOVIN P2Y PURINOCEPTOR 1 (ATP RECEPTOR...	221	7e-57
gi 1082756 pir S54253 purinergic receptor - human >gi 798836 e...	217	1e-55
gi 2829680 sp P79928 P2Y8_XENLA P2Y PURINOCEPTOR 8 (P2Y8) >gi 1...	217	2e-55
gi 2707256 gb AAC60339.1 (AF031897) G protein coupled P2Y nucl...	215	7e-55

blast to dbEST:

gi 7921097 gb AW827323.1 AW827323 hm31f01.x1 NCI_CGAP_Thy4 Homo...	46	0.022
gi 2900654 gb AA826657.1 AA826657 of34e08.s1 NCI_CGAP_Kid6 Homo...	42	0.35

Expression information for modulatory use (from cDNA panel screening):

Human uterus
 Human Placenta
 Human Brain
 Human Fetal Brain

2/7

```

1 MNEPLDYLAN ASDFPDYAAA FGNCTDENIP LKMHYLPVIY GIIFLVGFPG
51 NAVVISTYIF KMRPWKSSTI IMNLNACTDL LYLTSLPFLI HYYASGENWI
101 FGDFMCKFIR FSHFHNLYSS ILFLTCFSIF RYCVIIHPMS CFSIHKTRCA
151 VVACAVVWII SLVAVIPMTF LITSTNRNTR SACLDLTSSD ELNTIKWYNL
201 IILTATTFCFLP LVIVTLCYTT IHTTLTHGLQ TDSCLKQKAR RLTILLLLAF
251 YVCFLPFHIL RVIRIESRLL SISCSIENQI HEAYIVSRPL AALNTFGNLL
301 LYVVVSDNFQ QAVCSTVRCK VSGNLEQAKK ISYSNNP

```

FEATURES:**Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN_GLYCOSYLATIONN-glycosylation site

Number of matches: 4

1	10-13	NASD
2	23-26	NCTD
3	176-179	NRTN
4	179-182	NRSA

-----[2]

PDOC00004 PS00004 CAMP_PHOSPHO_SITEcAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1	240-243	RRLT
2	329-332	KKIS

-----[3]

PDOC00005 PS00005 PKC_PHOSPHO_SITEProtein kinase C phosphorylation site

Number of matches: 4

1	175-177	TNR
2	178-180	TNR
3	175-177	TNR
4	178-180	TNR

-----[4]

PDOC00006 PS00006 CK2_PHOSPHO_SITECasein kinase II phosphorylation site

Number of matches: 2

1	187-190	TSSD
2	188-191	SSDE

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	33	53	1.690	Certain
2	75	95	0.979	Putative
3	111	131	1.423	Certain
4	154	174	2.274	Certain
5	200	220	2.019	Certain
6	241	261	1.875	Certain
7	288	308	0.757	Putative

3/7

BLAST Alignment to Top Hit:

```
>gi|6679193|ref|NP_032798.1| purinergic receptor P2Y, G-protein
coupled 1 >gi|1352694|sp|P49650|P2YR_MOUSE P2Y
PURINOCEPTOR 1 (ATP RECEPTOR) (P2Y1) (PURINERGIC
RECEPTOR) >gi|767871|gb|AAA91302.1| (U22829) P2Y
purinoceptor [Mus musculus] >gi|6013075|emb|CAB57317.1|
(AJ245636) P2Y1 receptor [Mus musculus]
Length = 373
```

Score = 226 bits (571), Expect = 2e-58
 Identities = 109/299 (36%), Positives = 176/299 (58%), Gaps = 4/299 (1%)

```
Query: 24 CTDENIPLKMHYLPIYGIIFLVGFPGNAVVISTYIFKMRPWSSTIIMLNLAETDLYL 83
       C + +YLP +Y ++F++GF GN+V I ++F M+PW ++ M NLA D LY+
Sbjct: 42 CALTKTGFQFYLPAVYILVFIIGFLGNSVAIWMFVFHMKPWSGISVYMFNLALADFLYV 101

Query: 84 TSLPFLIHYYASGENWIFGDFMCKFIRFSFHNLSSSILFLTCFSIFRYCVIHPMSCFS 143
       +LP LI YY + +WIFGD MCK RF FH NLY SILFLTC S RY +++P+
Sbjct: 102 LTLPALIFYYFNKTDWIFGDAMCKLQRFIFHVNLYGSILFLTCISAHRYSGVVYPLKSLG 161

Query: 144 IHKTRCAVVACAVVWIISLAVIPMTFLITSTNRTNRS-ACLDLTSSDELNTIKWYNLIL 202
       K + A+ +VW+I +VA+ P+ F + R N++ C D TS+D L + Y++
Sbjct: 162 RLKKKNAIYVSVLVWLIVVVAISPILFYSGTGTTRKNKTVTCYDTSNDYLSYIFIYSMCT 221

Query: 203 TATTFCPLVIVTLCYTTIINTLTHGLQTDSCCLKQKARRLTLLLAFYVCFLPFHILRV 262
       T FC+PLV++ CY I+ L + +S L++K+ L I++L F V ++PFH+ ++
Sbjct: 222 TVAMFCIPLVLILGCGYGLIVKALIYNDLDNSPLRRKSIYLVIIVLTVFAVSYIPFHVMKT 281

Query: 263 IRIESRL---LSISCSIENQIHEAYIVSRPLAALNTFGNLLLYVVVSDNFQQAVCSTVR 318
       + + +RL C + + + Y V+R I A+LN+ + +LY + D F++ + R
Sbjct: 282 MNLRARLDFQTPEMCFNDRVYATYQVTRGLASLNCSVDPILYFLAGDTFRRRLSRATR 340
```

Hmmer search results (Pfam):

Scores for sequence family classification (score includes all domains):

N	Model	Description	Score	E-value
1	CE00175	CE00175 Purinoceptors	247.2	2.2e-70
1	PF00001	7 transmembrane receptor (rhodopsin family)	168.6	2.1e-52
1	CE00340	E00340 thrombin_receptor	101.2	1.4e-27
1	CE00530	CE00530 CHEMOKINE_RECEPTOR_TYPE_4	84.5	3.9e-26
2	CE00521	CE00521 THROMBIN_RECEPTOR	80.2	4.3e-28
2	CE00501	CE00501 ANGIOTENSIN_RECEPTOR	74.1	2e-19
2	CE00529	CE00529 CHEMOKINE_RECEPTOR_TYPE_3	62.2	1.4e-18
2	CE00505	CE00505 CHEMOKINE_RECEPTOR	54.9	1.3e-14
1	CE00319	E00319 adrenomedullin_receptor	49.7	1.6e-14
1	CE00531	CE00531 CHEMOKINE_RECEPTOR_TYPE_5	48.8	5.2e-13
1	CE00519	CE00519 BRADYKININ_RECEPTOR	48.8	2.2e-12
1	CE00207	CE00207 PURINERGIC	47.9	7.2e-14
1	CE00179	CE00179 Burkitts_lymphoma_receptor	39.1	5.6e-10

4/7

1 GTCCAGTGTG CTCGAAAGTG TCAGAATTTC CTTCATTTT AAAACTGAAT
 51 AATGTTCCAT GGAATGTATA CATCACATT TCTTCATCCA TTGATAGGCA
 101 CTTGGGCTGC ATCCACATT TGGCTATTGT GAATAATGCT GCTATGAACA
 151 TGGGTGTGCA AGTATCGTT TGAGTCTCAG CTTTCAATTTC TTTGGGTAT
 201 ATTCCCAGAA GCAGAATTGC TGGCTGTGT CATAATTCTA TTTTAATTTC
 251 TTAGAGGAAC CATTATACTT TTTTTTCTC ATAGCACCTG CACAATTTC
 301 ACTTACAGG ACTGGATATT TTTAAATGTC CTTTCCATGT CTAGAGTTCT
 351 AGGACCCAGT ACTTGTTTT CAAATAACCT CTTTAAGTCA TTTCTCGTT
 401 AGCTGGAAAT CTCGCTCACC ATGGAAAAAC ACTGGGATGT CTTTACTTC
 451 AGAAAGCCTT CAATTGGGTT TAGTCCCCCTC CCAGCCCCCTT ATTTACAAGT
 501 GGCAAGCCCT AATCATGTTT TTAATTATT TCTTGTCTT CTATAAGACA
 551 TCAGCTCTG GAAGCAAAGA TTATGTATTA TTCTTCATAA TCTCCACGGT
 601 GCTTGACACA GAATAATCAA ACAGCAAATG TCTGATCAAT GCATAGGTGG
 651 GAACAAAACA GGGTCAACTA TTTCCAGATG AAGAATTGAA AAAATGATTA
 701 AACATAGTAA ATTAGAGAA CATTTTTCT CCTAATGCAA AATATTTC
 751 TGGAGAGGCT GGAATTCAA TAATCAAAT GATAAGCATG ATAATAAAAA
 801 TGCCTAACTA GGGTATCCC AACCATCAGC CAGGGCGGAA TCTTCCACC
 851 TGGGCCCTGG TGGGACACCT AACCTAACAT TCTTGCTCCT GCTGGCTGGC
 901 ATTCCCAGAG GATGGGCTAT ACATGGTGT TGATGCCCAT CAACATGGAA
 951 GTGAGAAGA GACATTCAAT GCCCCGTCCAT GGAAGAGCGA AGAGGGCAT
 1001 TCTCAAGCA CCCCTATTG CTAATACAGT GAACCTTTG ATGCTGGAAG
 1051 AGATCTGAA GTTCAACGGT TCGGGAGGCG CTAGTATACT GGACCGATGTG
 1101 GCTGAGGTTG AGAGAAGAGA AAAGGCATT CCTAAGTTAT GTAGCAAGAA
 1151 GGCAACCTAC TCCCTAGATC TGGAGATACA GACAAACGCT CTTGCAAGGG
 1201 ATAGAAACTA TTTTTAAGCC ACCTTGCTAA GCGCTGGAGG GCAGGGTTGT
 1251 ATTGGATAC AGGGAAATGA CAACGCTTCA GAAAGTCGCC TCCATCCCAC
 1301 GCTAAATAAG AGCCACTGGG CGCCCTGGCC ACAGAGCGAG ACTCTTCTC
 1351 AAAAGAAAAA AAAAAAAAG AGGCACTGGG CTGTTAGGAGA GTGGTGAAT
 1401 TTCTGACCA AGACAAGGCG GCCACAACCC TCCCTGGGCA ACAACCCCTCC
 1451 CGCCGGGCTG GTTTTTGGG ACCTCTGCAA ACACCTTCTC CTGGCTCAGT
 1501 TCCCAGGCAA ACCGGGATGG AGGATGTCCA GAGGGTCCCAG GAGTGGAAAGC
 1551 CGCTGCTCCCT GTGTCTACTG GGAAGTGGGC GAGGGGGGAT GCTGCTGGCC
 1601 AGGCTGCTCG GGGCGAGGAC CCGAGAACAC GCACTTGCTC GCTGGGCTC
 1651 GGAGGGGTGG GAGGGGGCGGG GCGGTGGCGA GCCTCCCCGG GACAGGTCTG
 1701 CGGTCCGGGA CCAAATCAGGC TGACAGGCA CGCGGGGCGC ATGTCGGCCT
 1751 CGCCGGGGCTC CCAAGGTGAG CGCCGGGTGA TCGGGGAGGG GGGCCACCG
 1801 CGCCGGGGGA GCGGGTATAG GGGACCCCTGC CCTGACCCGG AGGGAGACGG
 1851 GAGGTGCCCCG AGGAGCAGGG GAAAGAGCAC CCCCACAGT CCCTGCAGCA
 1901 GGCAGGGTCG GACGGAGCAG CGCCCCCTTC TTCTCTGGAG GCAGCAGGAG
 1951 CACCCAGAT CCTGTGGCGG GAGGCGAAGT TTCAATTAG GGAACAAAAAA
 2001 GCCCAGGGCG AGAAGACATT TGACTTGAG AGAACTGGGG GAGTTAGGGA
 2051 TCGCTGATTT CATCTCACAC TTTTTTCTC ACCCAAGCGA GGGGTCTGGG
 2101 AGGTGGGCAG GAGAGTGTG TAGGAGACAA ACAGAGTTAA TCCCTCACCC
 2151 CATTGATCAC CGCCCTGCTG AATGTTCTCA AAGCTCACCC TGCACTGTGG
 2201 TGGACTGTGT GGTTTGGAA ACGACCAAA ACTCTGGCCC ACCCCCCACCC
 2251 CCAACAAGCG AGAAAATGT TTCCCCCTGAA ATGTCCTTCT TCTCTGTCTT
 2301 GTTTTTGTT TCGTTGCTAT GGATTCGTGA ACTAGCTTGT AATGGTCTAT
 2351 AAACCTCTAA CTTGTATGTA TAATATTGTC GTGTCCATGC ATGCATTTTT
 2401 CTGAAGCGAG CTCTCAGAAG TTGTATTGT TTCTCCGAA GAATAAAATAG
 2451 GCAAAACCAT ATCTTGGC TGCACTAACA TGACCTCTTA ATAGTCTTCA
 2501 GGGAAACTGA GTTGCAGATT CTTATTTAAC GAATTGGGA TTGGCCGGGT
 2551 GTGTGGCTCA CGTCTGTAAAT CCCAGCACTT TGGGAGGCTG AGGCAGGTGG
 2601 ATCACCTGAG GTCACTAGTT CGAGACCAGC CTGGCCAACA GGCTGGTGA
 2651 ACCCTGTCTC TATACAAATA CAAAAAAATTA GCCAGACGTG GTGGAACGGG
 2701 CCTGTAATCC CAGCCACTCC GAGGCTGAGG CAGGAGGATC CCTTGAACCC
 2751 AGGAGGTGGA GTTGAAGTG CGCCAAGATC ATGGCCACTG CACTCCAGTC
 2801 TGGGGACAG AGCAAGATT CGTCTCCAAA ACAACAAACAA CAATAATAAC
 2851 ACAAAACAAAC AAAAAACAAT AAAGTGGGA TTGAGGGAAG AGCCTTGAGG
 2901 CTCCAGAAG GTTGGGGAG AAAGATTCTC TCAGGGACCA GGGAGGACAG
 2951 CACTTACCCA GTCACAAATT TACCTGTGTT AAGAATTGGT CTCTTTCCA
 3001 CCAGTATGAA GTACAATTAG TTTTCAATAA TGAGACTGG GAAATCGAAG
 3051 TAAGTCTTG TGAATGAAAG GTCTACTTAA GTTTCTTTT GACAACCCAA
 3101 AGTGAACCCA GCTTCGGAGC TTCCAAGGAC TCTGGAGAGA AGGTAACAGA
 3151 TTAAGTGTTC CAACTGTGCA CCTCCCTCGT CCATGGCAT CTTGGAAAAA
 3201 GGAGGAAAGA ACCAACTGTG ACCTCAGCAT GGGAGGAGAA CCAGAAAAAC
 3251 CCAGGGAGTC TGGGGAGAG GGAGGAGAGC AGGGAGGAGGT ACAGGCATTG
 3301 TTGGATAATG AGCTTAACCTT ATCCAACAA GCTGATCTGC AAAACCAACA
 3351 GGCACCTTAG GAAAATCACA TTTCATCATT AATCCACCT TCTTGTTC

5/7

3401 CAAATGGGCA AGAGTACTCT GTGGAGGTGA AGTAGTATCT GCTGCCACTT
 3451 AACTGGATAG CATTCTGAT CCTGCCCTCT TTACCCACAG AATCTTGGAA
 3501 TCCCAATCCG TGAGGTTCTC GGGTGTGCTG GCATCAGGAC AGCGGTCCAC
 3551 GAACGGTGTG TTACCCAAT ATTGACATCC TGCAGCTGGC CTCAAACAAT
 3601 CACAGCTACT TTCCAATTTC AGAGAAAAAA AGGCTAAAAT TGTAAAGCTTT
 3651 TTCTTAACAA TTTTTTTATT TCTTAATTTC TGTGGATACA TAGCAGGTAT
 3701 ATATAATTAT GGGGTACATG AGATATTTC ATACAGGCAT GCAATGTGAA
 3751 ATAAGCACAT CATAGAGAAT GGGGTATCCA TCCACTCGAG CATGCATCCT
 3801 TTGTGTTATT AATACAAACA ATCCACTAC CTTCTTTATT TAAAAATGTG
 3851 CAATTAAGTT ATTATGACTA TGGTCACCCC ATTGTGCCAT CAAATAGTAG
 3901 GTCTTATTCA TTCTTTCTAC CTTCTTTTT TTTTTGTAC ACATTAGCCA
 3951 TCCCCCTCCAC CAGTCCCCCA CTACTCTTCC CAGCTCTGG TAACCCTCCT
 4001 TCTACTCTCT ATGCCCATGA GTTCAGFTGT TTTGATTTTT AGATCCCCATG
 4051 AATGAGTGAG AACATGCAAT GTTGTCTT CCATGCCCTGG CTTTTTCAC
 4101 TTCCCATAAT GACCTCCAGT TCTACCCATG AGGTTGCAAA TGACAGGATC
 4151 TCATTCTTT TTATGGCTGA ATAGTACTCC ACCATGTATA AGTACACAT
 4201 TTTCTTATC CATTCACTG TGGATGGACA CTTAGGTTGC TTCCAAATCT
 4251 TGGCTACTGT GAACAGGGCT GCAACAAACA TGGGAATGCG GATAGCTCTT
 4301 TGATATACTG ATTTCCTTT TTTTGGGTAT ATACTCAGCA GTGGAAATGC
 4351 TGGATCATAT GTTAGCTCAG TTTTTAGTT GCTGAGGAAC CTCCAAACTG
 4401 TTCTCCATAG TGATTTGACT AAAATTATAA GCTTTTTCTT CACAAGTTA
 4451 TTCCAACCTT TGATGCTGGC TTTTATTCAAT GCTGAACCTA GCTTAGTTCA
 4501 CACTATATT TTTGGTTATG TTAAAGCTGA ATACATCATT ATTTCTTGCT
 4551 TGCAAAGAAC CCATTGGTT CAAGTGATTT GTAACCTCCT CACGTTTGAG
 4601 AACTACTGTT GTAGAAAGAG AAGGGAATAT ACCCAGGGTT GTGCCACATT
 4651 TGTCTTTAAT GTCTTTGAG TTGAAATGAT TTGTTGTGG ACCATCTGGT
 4701 GTTCTGAGT GTGGCACTGA ATTGCACTGG GCCCTGAAGC TAGGGAGGAA
 4751 AGAAACAGCT TCATCCTACC TGCTCACAGC CCAGACCGGT TTTTGTGTG
 4801 CTGTGTTAA CATCACACGT ATTTTGTAAT CTCATACTTT TATGAGGTTT
 4851 TATGGAACAC AGATTAAGCT TCCTAGTCAC CCCCTCCACC CTTACAATTG
 4901 GCTAGTCAGT CTTCAATTCA TCTGAATGAA ACTATTGTTT TTCAGAACAA
 4951 TAAGGCAGTT GTTCTACCT ATAACCCATC TTTTCTTCAT AGAATTATTG
 5001 TTTAACATGG TCTCCATAAC ATGGCTAAGG GTTTATTGGA CAATGACAGG
 5051 AATTGTAATT TTGCTTCTTA CTTTAAAGT CATGAAAGAT CTAAGCTTGC
 5101 TTTTTTTTT TTTTTTTTT TGAAAGACTC TTGCTGTGTC ACTCAGGCTG
 5151 TAGTACAGTG GCTCAATCTC GGCTTACTCG AGCCTACACC TCCCAGGTT
 5201 AAGCGATTCT CGTGCCTCAG CCTCCCAAGT AGCTGGGACT ATAGGCATGG
 5251 GCCACACAC CCAGCTACGT TTTTTTTGT ATTTTTAATA GAGACGGAGT
 5301 TTTGCAATGC TGGCCAGGCA GGTCTCGAAC TCCTGACCTT GGGCGATCCG
 5351 CCTGCCCTCAG CCTCCCAAAG TGCTGAGGT ACAGGCTTGA GCCACTGTGC
 5401 CTGGCCTAAG AACCAGCTT TTTTTAGAA TGATAGTTGC CCCCTCAGTA
 5451 GCCCTTCAGT ACCATATAAG TTTTGCTGAT GTTACTGGTT GTTCTTTAAT
 5501 GCATTCAGCA AATATTAATA TTTACTGCAC ACCTTCCATG TGCCAGCCCC
 5551 TGGGGTATGA GGTGACTTGT GTTTTCTTT CAAGGAGCTG ACAAAATAAGC
 5601 AGCTGGACT TGAATTCCCT CCCCTTCCTC TCCTCTCAGC TCTCTGTTGA
 5651 GGCAATCCC AATTGGAATG AAATAACTCT GGAGAATTAT TTGGGGTT
 5701 ACTAAAACCA TCTTTACTTC TTTGCTCAG AAGCAACATT AGCTACTTAG
 5751 TGGCCTCTTG ACATAATCTA AGAACTGTCT TTGGGACGTG CTGCTTCTGG
 5801 GACTCCTTGG CTGTGACCAT TTTCTTCAC GTTTAAGTGA CTTCTGCTC
 5851 CTTATGCATG AGTCCTATTAA AGACATTCTG CCAAGGGAGA AGCCATCCAC
 5901 TGTGGTTGCC TGGCTTCAGA GCACTCTCAC TGGTCAGCAG TCCCTGACTC
 5951 CTATGACCTC CAGTGTGCTG TAGATGATAG TTTCTGAAT AGCACTTTA
 6001 ACATCTATGT TGCAATTAGT ATTATAAGT ATCTAGAGAT GATTTCACAGG
 6051 AGGATGGGCA TAGGTTAAAGT GTAAATACCA TACCAATTAA TTTTTATATC
 6101 AGTGACGTGA GCATCTGAG ATTGTTGGTAT CCTTGGGGGT CTGGGAACCA
 6151 ATGCACTGCA GATACTAAGG GACAACCTCA CAGTGTGTTGA TTTATTGAGC
 6201 TTTCTGAGAT ATTCACTCCTT TTCTGCTCTT CACACTGTT TTTCCCTCTA
 6251 TTATTTAGGG TAATCCTGAT GAAAATCAAC AAAATACACA TGAAGAGACA
 6301 GCACTGAGAG GTAAATGTGCT AACTGGCTT TGTGATTACA GGCACCTCAG
 6351 ATTTCCCTCT TGCTACTCA TTGTGAGATC CAGGCCAGCC CTGCAACAAAG
 6401 ACTGCCACAG TCGTATCTG GCTCACCCCT AACTAGCGGT TTGACTTTGA
 6451 GCACATCACT GCCTCCCTCG TGCCCTAGTT TCCTCTCTG TTGAATGGGG
 6501 ATGATCACAG CACTACCTCA CCATTGTTG TGAGGATTAA ATAGATTCAA
 6551 ATATGTAAGA CACTTACAGT AATGCTTGT ACACAGAAGG CACTATTATT
 6601 TTTTATTAAT CCCTATTTC CTTCCCAAC TTCACTCTCC AATATCCCAT
 6651 AACCATGCTG ATTTCCCTAC AACTCCCCA GACCTCCTGG ACTAAGTGAG
 6701 ATTTGGAAGA GTATACTCAG GGCAGTCAG AAGGACTGAT TCTGCTAATT
 6751 TAGGATTGTG CAAGTGGGT TAGTTGAAA TATACCTCTG TACTCCTCTC

6/7

6801 AGGAAATGCT GGATAGTGTG GATTTAGTAT CAGATAATAG TAATAATGGT
 6851 GGCAGACTCT GATGTTTCCT CCTTCCCTG TGATCTCAGT AGTCTTGTAC
 6901 CATTACATACA TCAAATGACT TAGCAACTAA CAATCACGTA ATATTTTTA
 6951 AGCTTTCTT ATGTGCCAG TAGTGAGGCC AGAGTTTTT TTCTTATTT
 7001 TAACATACTA TTGACAGCTC CATGAGGTGG GTTTATTTT TTGTTAAAATA
 7051 ATTTGCCATG TCCACTTCCC CAGTAGACTG TGACAGTGGC ATATACATCT
 7101 GTGTGTGTTA TTCAGTTGT ATCTCTAGGA CTTGACAGAG GACCCAAATAC
 7151 AGAGTGGGTG CTCCAGAAAT GTGGACTTGG ATGCATTCA CAAACAGCTA
 7201 TTTACATCCA AATAATAAGC TAATTATAAC TCTTCTTCTA GATGACTTT
 7251 TCTCATTGCA AAAATTCAAC GTCAAAATT TTAAAAAACAC TAATAAGCAA
 7301 GAAGAAGGAA GTTAAATTG CACTCCTTCC AGCCCTAAAG AGCTTAATCA
 7351 ACTTGCTTAG GAACACGCTA CTAATGGCT AATCTGAGCA CCAAATCTG
 7401 GCCATCTGGC TTCAGGGCTC AGCCTCTAA TGGCCGCATT ATACCATG
 7451 TCCCAGCCCCA TCCACCCAA CACACACACT CCCCTGCACA AACAAATTAT
 7501 GTTAGCTATT AAAAGGAATT GGAAGAGTGG AAATGTGCAA TTAAATCTAC
 7551 CACAAATACA ATGTTGCAAT ACGTGCAATC TATATTTTA AAATTACAG
 7601 TATTCAAGCAT TTTCTGTCT TCACACTTGT TCCCTCCGTT ATTTACGGCA
 7651 ATCATGACGA AAATCAACAA AATACAGATT GAGAGGCAGC ACTGAGAGGT
 7701 AATTAGAAAA GGGATGGCT TTTGAGATGAG AAAATGCGA GTGTGGATCC
 7751 CAGCTAAGCC CCTAGATGTT TTACAGAGA CGAGCCGCAA AACATTTCTC
 7801 AGTCTTAGTC TCTTGTAAAA TAGAGGTTAA AAGAAACACT TTTCAGTATT
 7851 TTGTGACATG TAGAAGTAAG TGATGGTGGC ATGCATCACA CTTGGTTAAT
 7901 AGTAGGTCTT GTTGTAACT CTCTAATGGC GATACCTAT GGCTTCTCCA
 7951 AATGGTGACC TTGCCAAATT GTTTCCAAA GCGACATGTG GCTTTTTCT
 8001 CCCAATCCCT CATTTTAACT CTCATGGTA TTTAACCTTT ATATTTTTAT
 8051 TAGATGCATT TAGTAACITG CCTCATAGTC ATTTCTTGG AAATTCAATT
 8101 TCTTCTCCAC AGGGTCTCTT TTGAGATTA AGAGAGAGAA GTGGCAAAATT
 8151 TAGGATGTTA GAATAATTTT CATTAAAAG TAGATCCTTG TTTTTATTAC
 8201 CCTATCATTA ATGTTTCTG TTTTCTTCA TCAGCGAGTT ACTGCTCATT
 8251 TGATTCATAT TGCCAAACTG AACTCTCTG TTTTCTTGCA AGATGAAAGG
 8301 AGACAACCAT GAATGAGCCA CTAGACTATT TAGCAAATGC TTCTGATTT
 8351 CCCGATTATG CAGCTGCTTT TGGAAATTGC ACTGATGAAA ACATCCCCT
 8401 CAAGATGCAC TACCTCCCTG TTATTTATGG CATTATCTTC CTCGTGGGAT
 8451 TTCCAGGCAA TGCAGTGTG ATATCCACTT ACATTTCAA AATGAGACCT
 8501 TGGAAAGAGCA GCACCATCAT TATGCTGAAC CTGGCTGCA CAGATCTGCT
 8551 GTATCTGACC AGCCTCCCT TCCGTATTCA CTACTATGCC AGTGGCGAAA
 8601 ACTGGATCTT TGGAGATTTC ATGTGTAAGT TTATCCGCTT CAGCTTCCAT
 8651 TTCAACCTGT ATAGCAGCAT CCTCTTCCTC ACCTGTTCA GCATCTCCG
 8701 CTACTGTGTG ATCATTCACC CAATGAGCTG CTTTTCCATT CACAAAACCTC
 8751 GATGTGCAGT TGTAGCCTGT GCTGTGGTGT GGATCATTTC ACTGGTAGCT
 8801 GTCATTCCGA TGACCTCTT GATCACATCA ACCAACAGGA CCAACAGATC
 8851 AGCCTGTCTC GACCTCACCA GTTCGGATGA ACTCAATACT ATTAAGTGGT
 8901 ACAACCTGTAT TTTGACTGCA ACTACTTTCT GCCTCCCTT GGTGATAGTG
 8951 ACATCTTGCT ATACCAACGAT TATCCACACT CTGACCCATG GACTGCAAAC
 9001 TGACAGCTGC TTAAACAGA AAGCAGGAAG GCTAACCTT CTGCTACTCC
 9051 TTGCATTTTA CGTATGTTT TTACCTTCC ATATCTTGAG GGTCAATTGG
 9101 ATCGAATCTC GCCTGTTTC AATCAGTTGT TCCATGAGA ATCAGATCCA
 9151 TGAAGCTTAC ATCGTTCTA GACCATTAGC TGCTCTGAAC ACCTTTGGTA
 9201 ACCTGTTACT ATATGTGGTG GTCAGCGACA ACTTCAGCA GGCTGTCTGC
 9251 TCAACAGTGA GATGCAAAGT AAGCGGGAAAC CTTGAGCAAG CAAAGAAAAT
 9301 TAGTACTCA ACAACCCCTT GAAATTTTC ATTTACTTAA CCAAAAACAA
 9351 ATACTTGCTG ATACTTTACCT TAGCATCTTA AGATGTTCA GATGTCTCCC
 9401 TCAATGGAAC TCCGTGTTAA TACTGTGTAT TCAAGTAATC ATGTGCCAA
 9451 GCCAGGGCAG AGCTTCTAGT TCTTGTCAAT CCCTTATTG AGCTCTCCA
 9501 CTGGGGAGAT ATAAGATGG GATGCATGTA TATCAGCAAA GTATTCAGAC
 9551 ATAGTATTAC AAGCTATTGG AACTCAGAGG CATCTTAGAG AACATCTGTT
 9601 CCCACCAACT TACTATATAT ACACGGAAAC CAATTCTTA CCCTTGCCCT
 9651 AGATTGCTCA GTAAATTGT GCCAAGATAG GAGAAAACCA ATCTTTTCAC
 9701 TCATCATTTC ATGCTTCTCT GCACTCTGG CCTATTGTA TTGAACCATT
 9751 AGACAATTCA ACCACTACT TGATCTTCA TTAATTTTA TTTTTACAT
 9801 CTCAGAGCTC TACAATTGT TTCTTCAAG CTTAACCTTG AGATTATAAA
 9851 ACTGGGTTTA GCCAGTTCTG TATATTACTT CAAGCCAGTA AGATACCCCT
 9901 GAAAT

FEATURES:

Start: 8309

Exon: 8309-9322

7/7

Stop: 9320

Map position for allelic variation:

11000292760378	#	SHGCNAME	CHROM#	LOD_SCORE	DIST. (cRs)
	1	SHGC-68276	13	8.71	14
	2	SHGC-8118	13	7.82	15
	3	SHGC-4149	13	7	22

Allelic variation (SNPs):

Position	Allele 1	Allele 2	Context
7,732	a	c	agaggcagcactgagaggtaatgtagaaggaaatggctttgagtcaga [a/c]aaa tgcgagtgtggatcccagctaagcccctagatgtgttaccagagacg
7,338	a	g	cactaataagcaagaagaaggtaaaattgcactccttccagcccta [a/g]aga gcttaatcaacttgcttaggaacacgctactaaatggctaattctgag
6,613	c	t	cttacagtaatgctttgtacacagaaggcactattatttttatataatcc [c/t]tat tttccttcccaacttcatctccaaatatcccataaccatgctgatt

				Protein		
POSITION	Allele 1	Allele 2		Position		
7,732	a	c	Intron			
7,338	a	g	Intron			
6,613	c	t	Intron			

SEQUENCE LISTING

<110> WEI, Ming-Hui et al

<120> ISOLATED HUMAN G-PROTEIN COUPLED
RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCRs
PROTEINS, AND USES THEREOF

<130> CL000750-PCT

<140> (tba)

<141> 2001-05-17

<150> US 09/634,656

<151> 2000-08-08

<150> US 60/205,196

•<151> 2000-05-18

<160> 7

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1014

<212> DNA

<213> Human

<400> 1

atgaatgagc	cactagacta	tttagcaaat	gcttctgatt	tcccccgatta	tgcagctgct	60
tttggaaatt	gcactgatga	aaacatccca	ctcaagatgc	actacccccc	tgttattttat	120
ggcattatct	tcctcggtgg	atttccaggg	aatgcagtag	tgatatccac	ttacattttc	180
aaaatgagac	cttggaaagag	cagcaccatc	attatgctga	acctggcctg	cacagatctg	240
ctgtatctga	ccagccccc	cttcctgatt	caactactatg	ccagtgccga	aaactggatc	300
tttggagatt	tcatgtgtaa	gtttatccgc	ttcagcttcc	atttcaacct	gtatagcagc	360
atccctctcc	tcaccctgttt	cagcatctc	cgctactgtg	tgatcattca	cccaatgagc	420
tgctttcca	ttcacaaaaac	tcgatgtgca	gtttagcct	gtgctgtgg	gtggatcatt	480
tcactggtag	ctgtcattcc	gatgacccctc	ttgatcacat	caaccaacag	gaccaacaga	540
tcagccctgc	tcgacacctac	cagttcggt	gaactcaata	ctattaagtg	gtacaacacgt	600
attttgcgt	caactacttt	ctggctcccc	ttggtgatag	tgacactttg	ctataccacg	660
attatccaca	ctctgaccca	tggactgcaa	actgacagct	gccttaagoa	gaaagcacga	720
aggctaaacc	ttctgtctact	ccttgcattt	tacgtatgtt	ttttaccctt	ccatatcttg	780
agggtcattc	ggatcgaatc	tcgcctgctt	tcaatcagtt	gttccattga	gaatcagatc	840
catgaagctt	acatcgttc	tagaccattt	gctgtctga	acaccttgg	taacctgtta	900
ctatatgtgg	tggtcagcga	caactttcag	caggctgtct	gctcaacagt	gagatgcaaa	960
gtaaaggcggg	accttgcgca	aqcaaaaqaaa	attagttact	caaacaaccc	ttga	1014

<210> 2

<211> 337

<212> PBT

<213> Human

<400> ?

```

Met Asn Glu Pro Leu Asp Tyr Leu Ala Asn Ala Ser Asp Phe Pro Asp
  1           5           10           15
Tyr Ala Ala Ala Phe Gly Asn Cys Thr Asp Glu Asn Ile Pro Leu Lys
  20          25          30
Met His Tyr Leu Pro Val Ile Tyr Gly Ile Ile Phe Leu Val Gly Phe
  35          40          45
Pro Gly Asn Ala Val Val Ile Ser Thr Tyr Ile Phe Lys Met Arg Pro
  50          55          60

```

Trp Lys Ser Ser Thr Ile Ile Met Leu Asn Leu Ala Cys Thr Asp Leu
 65 70 75 80
 Leu Tyr Leu Thr Ser Leu Pro Phe Leu Ile His Tyr Tyr Ala Ser Gly
 85 90 95
 Glu Asn Trp Ile Phe Gly Asp Phe Met Cys Lys Phe Ile Arg Phe Ser
 100 105 110
 Phe His Phe Asn Leu Tyr Ser Ser Ile Leu Phe Leu Thr Cys Phe Ser
 115 120 125
 Ile Phe Arg Tyr Cys Val Ile Ile His Pro Met Ser Cys Phe Ser Ile
 130 135 140
 His Lys Thr Arg Cys Ala Val Val Ala Cys Ala Val Val Trp Ile Ile
 145 150 155 160
 Ser Leu Val Ala Val Ile Pro Met Thr Phe Leu Ile Thr Ser Thr Asn
 165 170 175
 Arg Thr Asn Arg Ser Ala Cys Leu Asp Leu Thr Ser Ser Asp Glu Leu
 180 185 190
 Asn Thr Ile Lys Trp Tyr Asn Leu Ile Leu Thr Ala Thr Thr Phe Cys
 195 200 205
 Leu Pro Leu Val Ile Val Thr Leu Cys Tyr Thr Thr Ile Ile His Thr
 210 215 220
 Leu Thr His Gly Leu Gln Thr Asp Ser Cys Leu Lys Gln Lys Ala Arg
 225 230 235 240
 Arg Leu Thr Ile Leu Leu Leu Ala Phe Tyr Val Cys Phe Leu Pro
 245 250 255
 Phe His Ile Leu Arg Val Ile Arg Ile Glu Ser Arg Leu Leu Ser Ile
 260 265 270
 Ser Cys Ser Ile Glu Asn Gln Ile His Glu Ala Tyr Ile Val Ser Arg
 275 280 285
 Pro Leu Ala Ala Leu Asn Thr Phe Gly Asn Leu Leu Leu Tyr Val Val
 290 295 300
 Val Ser Asp Asn Phe Gln Gln Ala Val Cys Ser Thr Val Arg Cys Lys
 305 310 315 320
 Val Ser Gly Asn Leu Glu Gln Ala Lys Lys Ile Ser Tyr Ser Asn Asn
 325 330 335
 Pro

<210> 3
 <211> 9905
 <212> DNA
 <213> Human

<400> 3

 gtccagtgtc ctcgaaatgt tcagaatttc cttcattttt aaaactgaat aatgttccat 60
 ggaatgtata catcacattt tcttcatcca ttgataggca ctggggctgc atccacattt 120
 tggctattgt gaataatgtc gctatgaaca tgggtgtgca agtatccgt tgagtctcg 180
 ctttcaattt tttgggtat attccagaa gcagaattgc tggcttggtt cataattcta 240
 ttttaattt ttagaggaac cattatactt tttttttcc atagcacctg cacaatttta 300
 actttacagg actggatatt tttaatgtc cttccatgt ctagagttct aggacccagt 360
 acttggtttt caaataaacct cttaagtca ttccctcggt agctggaaat cctgctcacc 420
 atggaaaaac actggatgt ctttacttc agaaaacccctt cattgggtt tagccccctc 480
 ccagccccctt attacaagt ggcaagccct aatcatgttt ttaatttttatt tcttggtttc 540
 ctataagaca tcagctctg gaagcaaaga ttatgttata ttcttcaata tctccacgg 600
 gcttgacaca gaataatcaa acagcaaattt tctgtatcaat gcataggtgg gaacaaaaca 660
 gggtaacta ttccagatg aagaattgaa aaaaatgatta aacatagtaa aatttagagaa 720
 cattttttct cctaattgtca aatattttcc tggagaggct ggaattcaaa taatcaaaat 780
 gataaggatg ataataaaaaa tgccttaacta gggctatccc aaccatcagc cagggcggaa 840
 tctttccacc tggggccctgg tgggacaccc ttttgcctt gctggctggc 900
 atccccagag gatgggttat acatgggtt tgatgcccattt caacatggaa gtgcagaaga 960
 gacatcaat gcccgtccat ggaagagcga agagggggcat tctccaagca cccccattcg 1020
 ctaatacagt gaactctttg atgcttggaa agatcttggaa gttcaacgggt tcgggaggcg 1080
 ctatgtataact ggacgatgtc gctgagggttc agagaagaga aaaggcgatt ccttaagttat 1140

tctgaattga actattgttt ttcagaacaa taaggcagtt gtttctacct ataacccatc 4980
 ttttcttcat agaattattt ttaacatgg tctccataac atggctaagg gtttatttgg 5040
 caatgacagg aattgttaatt ttgcttccta cttaaaagt catgaaagat ctaagcttgc 5100
 tttttttttt tttttttttt tgaaagagtc ttgtgtgtc actcaggctg tagtacagt 5160
 gtcatatctc ggttactgc agcctacacc tcccagggtc aagcgattt cgtgcctcag 5220
 cttcccaagt agctgggact ataggcatgg gccaccacac ccagctacgt tttttttgt 5280
 attttaataa gagacggagt ttgcaatgc tggccaggca ggtctcgaac tcctgacctt 5340
 gggcgatccg cctgcctcag cctcccaaag tgctgaggtt acaggctga gccactgtgc 5400
 ctggcctaag aaccaagctt ttttttagaa tgatagttgc cccttcagta gcccttcagt 5460
 accatataag ttttgcgtat gttactgggtt gttcttaat gcattcagca aatattaata 5520
 tttaactgcac accttccatg tgccagcccc tgggttatga ggtgacttgg ttttttctt 5580
 caaggagctg acaaataaagc agcttgact tgaattccct ccccttcct tcctctcag 5640
 tctctgtga gccaaatccc aattggaaatg aaataactct ggagaattat tttggcggtt 5700
 actaaaacta tctttacttc ttgctccag aagcaacatt agctacttag tggcctctt 5760
 acataatcta agaactgtct ttgggacgtg ctgcttctgg gactccctgg ctgtgaccat 5820
 tttcttcaaa gtttaagtga ctccctgtc cttatgcattt agtccattt agacattctg 5880
 ccaaggaga agccatccac tgggttgcc tggcttcaga gcaactctcag tggcagcag 5940
 tccctgactc ctatgacctc cagtgctgt tagatgatag tttcatgaat agcacttta 6000
 acatctatgt tgcatcttagt attataagta atctagagat gatttacagg aggatggca 6060
 taggttaagt gtaaatacca taccatttt tttttatatc agtgacgtga gcatctgcag 6120
 attttggat ccttgggggt ctgggaacca atgactgca gatactaagg gacaactcta 6180
 cagtgtttga ttatttggc tttctgagat attcagcctt ttccctgtt cacacttgg 6240
 ttccctcta ttatttaggg taatcctgat gaaaatcaac aaaatacaca tgaagagaca 6300
 gcactgagag gtaatgtgct aactggctt tggattaca ggcacttcag atttccctt 6360
 tgcctactca ttgtgagatg cagcccaagcc ctgcaacaag actgccacag tcgtatcctt 6420
 gtcacccct aactagcggt ttgacttga gcacatcaact gcttcctccg tgcctcagtt 6480
 tcctctctg ttaatgggg atgatcacag cactacctca ccatttggg tgaggattaa 6540
 atagattcaa atatgtaaag cacttacagt aatgtttgtt acacagaagg cactattatt 6600
 ttttatttaat ccctattttt cttcccaac ttcatctccc aatatcccat aaccatgctg 6660
 atttccttac aactccccca gacctctgg actaaatggg agtggaaatgtacttgc 6720
 ggcagtcaag aaggactgat tctgtaatt tagattttt caagtgggt tagttgaaa 6780
 tataccctctg tactcctctc agggaaatgtt ggatagtgta gatttagt cagataatag 6840
 taataatgtt ggcagactct gatgtttcctt cttcccttg tcatctcagt agtcttgac 6900
 cattcataca tcaaataactt tagcaactaa caatcacgtt atgcatttca ccaacagcta 6960
 atgtgcccag tagtgagcca agatgtttt ttccattttt taacataacta ttgacagctc 7020
 catgaggtgg gtttatattt tggtaaaaat atttgccatg tccacttccc cagtagactg 7080
 tgacagtggc atatacatctt gttgtgttta ttcaattttt atctcttagga cttgacagag 7140
 gacccaataac agagtgggtg ctccagaaat gtggacttgg atgcatttca ccaacagcta 7200
 ttacatcca aataataaagc taatttataac tcttcttcta gatgtacttt tctcatttgc 7260
 aaaatttccacc gtcaaaattt taaaacac taataagcaa gaagaaggaa gttaaaattt 7320
 cactcctcc agccctaaag agcttaatca acttgcttag gaacacgcta ctaaatggct 7380
 aatctgagca cccaaatctg gccatctggc ttccaggcacttccctcataa tggccgcatt 7440
 ataccatttga tcccgccccca tccaccccaa cacacacact cccctgcaca aacaaatttat 7500
 gtttagtattt aaaaaggaaat ggaagagtgg aaatgtgcaat taaaatctac cacaatataca 7560
 atgttgcattt acgttgcattt tatattttttaa atatttacag tatttcgcattt tttcctgtct 7620
 tcacacttgc tccctccgtt atttacggca atcatgacga aatcaacaa aatacagatt 7680
 gagaggcagc actgagaggtt aatgttagaaa gggaaatggct tttgagtcag aaaaatgcga 7740
 gtgtggatcc cagctaaaggcc cctagatgtt ttaccagaga cgagccgaa aacatttctc 7800
 agtcttagtc tcttggaaaat taggttaat aaaaacact tttcagttt ttgtgacatg 7860
 tagaagtaag tgatgggttgc atgcatttca cttgggttaat agtaggttctt gttgttaagt 7920
 ctctaatggc gataccctat ggttcttca aatgtgtacc ttgccaaattt gtttccaaa 7980
 ggcacatgtt gctttttctt cccaaatccctt cattttactt ctcattgtttaa ttaactttt 8040
 atattttat tagatgcattt tagtaacttgc cctcatagtc attttcttgg aaattcaattt 8100
 tcttctccac agggcttctt ttgagattaa agagagagaa gtggccaaattt taggtatgtt 8160
 gaataattttt cattttaaaat tagatccttgc tttttatttttac ccttatttca atgtttctt 8220
 ttttccattt tcaagcgagttt actgtcttattt tgatttcatat tggccaaacttgc aactcttgc 8280
 ttttcttgc agatgaaagg agacaaccat gaatgagcca cttagactt tagcaaatgc 8340
 ttctgatttcc cccgattatg cagctgtttt tggaaatttgc actgtatggaa acatccact 8400
 caagatgcac tacctccctg ttatatttttgc cattatctt ctcgtggat ttccaggca 8460
 tgcagtagtgc atatccactt acattttcaaa aatgagacct tggaaagagca gcaccatcat 8520
 tatgtgttgc acatccactt acatccactt acatccactt acatccactt acatccactt 8580
 ctactatgcc agtggcgaaa actggatctt tggagatttca atgtgttaatg ttatccgctt 8640
 cagcttccat ttcaacccatgtt atgcacatccat ccttccatccat acctgtttca gcatcttccg 8700

ctactgtgtg atcattcacc caatgagctg cttttccatt cacaaaaactc gatgtgcagt 8760
 tgtagcctgt gctgtgggt ggatcatttc actggtagct gtcattccga tgaccttctt 8820
 gatcacatca accaacagga ccaacagatc agcctgtctc gacctcacca gttcggatga 8880
 actcaatact attaagtggt acaacctgtat tttgactgca actactttct gcctcccctt 8940
 ggtgatagtg acactttgtc ataccacat tatecacact ctgaccatc gactgcaaac 9000
 tgacagctgc cttaaagcaga aagcacgaag gctaaccatt ctgctactcc ttgcattttt 9060
 cgtatgtttt ttacccttcc atatcttggg ggtcattcgg atcgaatctc gcctgcttc 9120
 aatcaatgtt tccattgaga atcagatcca tgaagcttac atcggttctt gaccattagc 9180
 tgctctgaac acctttggta acctgttaat atatgtggt gtcagcgaca actttcagca 9240
 ggctgtctgc tcaacagtga gatgcaaagt aagcgggaac cttagcagaag caaagaaaat 9300
 tagttactca aacaaccctt gaaatatttc atttacttaa ccaaaaacaa atacttgctg 9360
 atacttacc tagcatccta agatgttcag gatgtctccc tcaatggaaac tcctggtaaa 9420
 tactgtgtat tcaagtaatc atgtgccaaa gcaggggcag agcttcttagt tctttcaat 9480
 cccttatttgc agctcctcca ctggggagat ataagaatgg gatgcatgta tattcagcaaa 9540
 gtattcagac atagtattac aagcttattgg aactcagagg catcttagag aacatctgtt 9600
 cccaccaact tactatataat acacggaaac caattctta cccttgcctt agattgctca 9660
 gtaaaatttgc gccaagatag gagaaaacca atctttcac tcattatc atgcttctct 9720
 gcactctggg cctatttgc ttgaaccatt agacaattca aaccactact tttatcttc 9780
 ttaatattta ttttttacat ctcagagctc tacaatttgc ttccattcaag cttaactttg 9840
 agattataaa actgggttta gccagttctg tatattactt caagccagta agataccctt 9900
 gaaat 9905

<210> 4
 <211> 299
 <212> PRT

<213> Mus musculus

<400> 4
 Cys Ala Leu Thr Lys Thr Gly Phe Gln Phe Tyr Tyr Leu Pro Ala Val
 1 5 10 15
 Tyr Ile Leu Val Phe Ile Ile Gly Phe Leu Gly Asn Ser Val Ala Ile
 20 25 30
 Trp Met Phe Val Phe His Met Lys Pro Trp Ser Gly Ile Ser Val Tyr
 35 40 45
 Met Phe Asn Leu Ala Leu Ala Asp Phe Leu Tyr Val Leu Thr Leu Pro
 50 55 60
 Ala Leu Ile Phe Tyr Tyr Phe Asn Lys Thr Asp Trp Ile Phe Gly Asp
 65 70 75 80
 Ala Met Cys Lys Leu Gln Arg Phe Ile Phe His Val Asn Leu Tyr Gly
 85 90 95
 Ser Ile Leu Phe Leu Thr Cys Ile Ser Ala His Arg Tyr Ser Gly Val
 100 105 110
 Val Tyr Pro Leu Lys Ser Leu Gly Arg Leu Lys Lys Asn Ala Ile
 115 120 125
 Tyr Val Ser Val Leu Val Trp Leu Ile Val Val Val Ala Ile Ser Pro
 130 135 140
 Ile Leu Phe Tyr Ser Gly Thr Gly Thr Arg Lys Asn Lys Thr Val Thr
 145 150 155 160
 Cys Tyr Asp Thr Thr Ser Asn Asp Tyr Leu Arg Ser Tyr Phe Ile Tyr
 165 170 175
 Ser Met Cys Thr Thr Val Ala Met Phe Cys Ile Pro Leu Val Leu Ile
 180 185 190
 Leu Gly Cys Tyr Gly Leu Ile Val Lys Ala Leu Ile Tyr Asn Asp Leu
 195 200 205
 Asp Asn Ser Pro Leu Arg Arg Lys Ser Ile Tyr Leu Val Ile Ile Val
 210 215 220
 Leu Thr Val Phe Ala Val Ser Tyr Ile Pro Phe His Val Met Lys Thr
 225 230 235 240
 Met Asn Leu Arg Ala Arg Leu Asp Phe Gln Thr Pro Glu Met Cys Asp
 245 250 255
 Phe Asn Asp Arg Val Tyr Ala Thr Tyr Gln Val Thr Arg Gly Leu Ala
 260 265 270

Ser Leu Asn Ser Cys Val Asp Pro Ile Leu Tyr Phe Leu Ala Gly Asp
275 280 285
Thr Phe Arg Arg Arg Leu Ser Arg Ala Thr Arg
290 295

<210> 5
<211> 101
<212> DNA
<213> Human

<400> 5
agaggcagca ctgagaggtta atgttagaaag ggaatggctt ttgagt~~caga~~ maaatgc~~gag~~ 60
tgtggatccc agctaagccc ctagatgtgt taccagagac g 101

<210> 6
<211> 101
<212> DNA
<213> Human

<400> 6
cactaataag caagaagaag gaagt~~aaaa~~ ttgcactcct tccagcc~~cta~~ ragagcttaa 60
tcaacttgct taggaacacg c~~tactaaatg~~ gctaatctga g 101

<210> 7
<211> 101
<212> DNA
<213> HUMAN

<400> 7
cttacagtaa tgct~~tt~~gtac acagaaggca ctattat~~ttt~~ ttattaatcc ytattttcc 60
ttcccaactt catctcccaa tatcccataa ccatgctgat t 101

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/087980 A3

(51) International Patent Classification⁷: C07K 14/705, C12N 15/12, C07K 16/28, A01K 67/027, A61K 38/17

(74) Agent: CELERA GENOMICS CORP.; Robert A. Millman, Patent Director, 45 West Gude Drive C2-4, Rockville, MD 20850 (US).

(21) International Application Number: PCT/US01/15957

(81) Designated States (national): AE, AG, AL, AM, AT, AU,

(22) International Filing Date: 17 May 2001 (17.05.2001)

AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,

(25) Filing Language: English

DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,

(26) Publication Language: English

HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,

(30) Priority Data:

LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,

60/205,196 18 May 2000 (18.05.2000) US

NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,

09/634,656 8 August 2000 (08.08.2000) US

TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(71) Applicant: APPLERA CORPORATION [US/US];
Robert A. Millman, Assistant Secretary, 761 Main Avenue,
Norwalk, CT 06859 (US).

(84) Designated States (regional): ARIPO patent (GH, GM,

(72) Inventors: WEI, Ming-Hui; Celera, 45 West Gude Drive,
Rockville, MD 20850 (US). ZHAO, Qi; Celera, 45 West
Gude Drive, Rockville, MD 20850 (US). CRAVCHIK,
Anibal; Celera, 45 West Gude Drive, Rockville, MD 20850
(US). DI FRANCESCO, Valentina; Celera, 45 West Gude
Drive, Rockville, MD 20850 (US). BEASLEY, Ellen M.;
Celera, 45 West Gude Drive, Rockville, MD 20850 (US).

KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:

18 July 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A3

WO 01/087980

(54) Title: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the Human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/15957A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/705 C12N15/12 C07K16/28 A01K67/027 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LEON C ET AL: "Cloning and sequencing of a human cDNA encoding endothelial P2Y1 purinoceptor" GENE, ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, NL, vol. 171, no. 2, 1 June 1996 (1996-06-01), pages 295-297, XP004042812 ISSN: 0378-1119 cited in the application the whole document ---- -/-	1,2,4,5, 8-11,13

 Further documents are listed in the continuation of box C. Patent family members are listed in annex.

* Special categories of cited documents :

- 'A' document defining the general state of the art which is not considered to be of particular relevance
- 'E' earlier document but published on or after the international filing date
- 'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- 'O' document referring to an oral disclosure, use, exhibition or other means
- 'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

14 March 2002

28/03/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Mateo Rosell, A.M.

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/US 01/15957

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	TOKUYAMA Y ET AL: "CLONING OF RAT AND MOUSE P2Y PURINOCEPTORS" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 211, no. 1, 6 June 1995 (1995-06-06), pages 211-218, XP000673449 ISSN: 0006-291X figure 1 ---	1,2,4,5, 8-11,13
A	WO 97 20045 A (COR THERAPEUTICS INC) 5 June 1997 (1997-06-05) page 1, line 20 -page 2, line 20; figures 1A-C; examples 1,2 ---	1,2,4,5, 8-11,13
A	VOLONTE C ET AL: "SELECTED P2 PURINOCEPTOR MODULATORS PREVENT GLUTAMATE-EVOKED CYTOTOXICITY IN CULTURED CEREBELLAR GRANULE NEURONS" JOURNAL OF NEUROSCIENCE RESEARCH, WILEY-LISS, US, vol. 45, no. 2, 15 July 1996 (1996-07-15), pages 183-193, XP002048687 ISSN: 0360-4012 the whole document ---	14,17
E	WO 01 36471 A (ARENA PHARMACEUTICALS INC ;CHEN RUOPING (US); DANG HUONG T (US); L) 25 May 2001 (2001-05-25) page 113-114 abstract page 3, line 10-13 page 13, line 1 -page 14, line 7; table B example 1; table C page 34, line 7-21 ---	1-17
E	WO 01 49847 A (WHITE DAVID ;GLUCKSMANN MARIA ALEXANDRA (US); MILLENNIUM PHARM INC) 12 July 2001 (2001-07-12) abstract page 1, line 18 -page 5, line 26; figure 1 ---	1-17
E	WO 01 87937 A (INCYTE GENOMICS INC ;KALLICK DEBORAH A (US); PATTERSON CHANDRA (US) 22 November 2001 (2001-11-22) SEQ.ID.N.10 abstract page 5, line 35 -page 10, line 25 ---	1-17
		-/-

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/US 01/15957

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	<p>DATABASE EMBL 'Online! 1 December 2001 (2001-12-01) Database accession no. Q96P68 XP002193122 G PROTEIN-COUPLED RECEPTOR. GPR80. Homo sapiens abstract -----</p>	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/15957

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9720045	A 05-06-1997	US 5871963 A		16-02-1999
		AU 729875 B2		15-02-2001
		AU 1075397 A		19-06-1997
		CA 2237937 A1		05-06-1997
		EP 0868510 A2		07-10-1998
		JP 2000500658 T		25-01-2000
		WO 9720045 A2		05-06-1997
		US 6063582 A		16-05-2000
-----	-----	-----	-----	-----
WO 0136471	A 25-05-2001	AU 1769601 A		30-05-2001
		WO 0136471 A2		25-05-2001
-----	-----	-----	-----	-----
WO 0149847	A 12-07-2001	AU 2458101 A		16-07-2001
		WO 0149847 A2		12-07-2001
-----	-----	-----	-----	-----
WO 0187937	A 22-11-2001	WO 0187937 A2		22-11-2001
		WO 0190359 A2		29-11-2001
-----	-----	-----	-----	-----